Improving model validation in health technology assessment

Vemer, Pepijn; Krabbe, P. F. M.; Van Voorn, G. A. K.; Ramos, Corro; Al, M. J.; Feenstra, T. L.

Published in:
Value in Health

DOI:
10.1016/j.jval.2013.06.015

Publication date:
2013

VALUE IN HEALTH 16 (2013) 1106–1107

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Improving Model Validation in Health Technology Assessment: Comments on Guidelines of the ISPOR-SMDM Modeling Good Research Practices Task Force

In 2012, a joint task force from ISPOR and SMDM produced a set of guidelines on modeling good practices for research [1]. These articles are highly influential in the field of health technology assessment, and rightly so. Therefore, we would like to discuss some concerns they raise about the definition and interpretation of an important aspect, namely, model validation [2].

Definition

The task force defined validation as "a set of methods for judging a model's accuracy in making relevant predictions." This use of the term "accuracy" goes back to definitions for the fields of engineering and computer sciences, where validation has a long tradition [3–5]. Crucially, the level of accuracy required in engineering models can be defined a priori and determined a posteriori. A benchmark usually exists (a drainage system, a dam, a motor) to which one can compare the outcomes. Health-economic decision models, in contrast, facilitate estimation of the consequences of health care decisions, using outcomes that do not lend themselves to benchmarking. Moreover, in models without human components, the dynamic behavior to be modeled follows laws of nature [6], and so they do not need to account for uncertainty. Models in health economics do, however.

This difference prompted us to adapt the definition of model validation for use in the health care setting by avoiding the term accuracy, reflecting the uncertainties intrinsic to our work. In line with the literature in other fields [6–8], validation could be defined as the act of evaluating whether a model is a proper and sufficient representation of the system it is intended to represent in view of an application. Here, "proper" means that the model is in accordance with what is known about the system and "sufficient" means that the results can serve as a solid basis for decision making.

The Validation Process

The task force distinguishes models to be used once (single-application) from those to be used several times (multiapplication). The task force states that validation should be ongoing for a multiapplication model, while it can be conducted once for a single-application model [2]. In our opinion, however, validation should be continuous [5,9], regardless of the number of times a model is applied. Unfortunately, this recommendation is seldom followed [4].

Conducting validation throughout the modeling process would be worth every effort because mistakes could then be found and corrected at an earlier stage. Validating once toward the end leaves little time to remedy any problems they produce. Furthermore, the likelihood of finding mistakes increases with the number of validation rounds, minimizing the chance that the model will contain serious errors when decisions are made on its basis. In addition, when stakeholders are continuously involved in building the model, instead of being asked for comments afterwards, they will be more inclined to accept it, because they understand its rationale. Altogether, more validation points could make model development shorter and cheaper.

While differentiating types of models for validation purposes is useful, the focus should lie on the outcomes and the decisions to be based on them, rather than on the number of times the model will be used. For example, verification examines the extent to which the mathematical calculations are performed correctly and are consistent with the model's specifications [2]. One of the most powerful verification tools is double coding, whereby some or all the code is programmed twice and independently. Although very time consuming, this yields a very high level of verification. Yet it is unlikely that the amount of time, effort, and money needed for double coding will be invested in a model with a relatively small impact, even if it is used more than once. Nonetheless, double coding may be cost-effective and perhaps even necessary for models supporting decisions with large budget impacts.

When Is a Model Valid Enough?

According to the authors of the guidelines, "it is not possible to specify criteria that a model must meet to be declared 'valid.' " This is true, though we would add that it is possible to operationalize a level of validity: a model can be "valid enough" to reliably support a decision to be based on its outcomes. In this respect, corroboration is a relevant term [10]. Passing any validation test corroborates the model and increases its credibility. One of the aims of a project we have just started is to operationalize validation guidelines by selecting tools on the grounds of their relevance. We would then add workable criteria, keeping in mind the impact of the decisions ultimately to be made on health care and budgets. A work in progress, we expect it to improve the model review process by providing an overview of validation tools applied during its development and their results in relation to the model's aim.

Source of financial support: This publication is part of a study funded by ZonMw, The Netherlands Organization for Health Research and Development (ZonMw project number 80-82500-98-12211).

Conclusions

We have proposed an alternative definition of validation because we felt that the definition proposed by the task force is not adequate for health-economic decision models. Any comments on our proposal are welcome. We have also recommended introducing validation as a continuous process in all types of models, thereby dropping the distinction between single-purpose and multiuse models. Finally, while agreeing with the task force that it may not be possible to declare a model "valid," we have added that it might be possible to call a model "valid enough" for its purpose. Introducing a continuum of validity is practically useful and should not be disregarded.

The ISPOR/SMDM Modeling Good Research Practices Task Force has assembled a state-of-the-art overview of health-economic decision modeling and has been rightly commended for this work. Any state of the art, however, is a moving target. Experience in other scientific fields is worth taking into account to help move the target in health-economic decision modeling, toward providing reliable support for the important policy decisions facing health care systems today.

Pepijn Vemer, P.F.M. Krabbe, T.L. Feenstra
Department of Epidemiology, University of Groningen/University Medical Center, Groningen, The Netherlands

G.A.K. Van Voorn
Biometris, Wageningen University & Research, The Netherlands

Corro Ramos, M.J. Al
Institute for Medical Technology Assessment, Erasmus University, Rotterdam, The Netherlands

1098-3015/$36.00 – see front matter
http://dx.doi.org/10.1016/j.jval.2013.06.015

REFERENCES

[1] Caro JJ, Briggs AH, Siebert U, Kuntz KM. ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value Health 2012;15:796–803.
[2] Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Value Health 2012;15:843–50.
[3] Schlesinger S, Crosbie RE, Gagné RE, et al. Terminology for model credibility. Simulation 1979;32:103–4.
[4] Peters B, Smith J, Medeiros D, Rohrer M, eds. How to build valid and credible simulation models. In: 2001 Winter Simulation Conference; 2001.
[5] Ingalls R, Rossetti M, Smith J, Peters B, eds. Validation and verification of simulation models. In: 2004 Winter Simulation Conference. Piscataway, NJ: IEEE Press, 2004.
[6] Kleijnen JPC. Ethical issues in engineering models: an operations researcher's reflections. Sci Eng Ethics 2011;17:539–52.
[7] Ljung L. System Identification: Theory for the User. Englewood Cliffs, NJ: PTR Prentice-Hall, 1987.
[8] Ngo TA, See LM. Calibration and validation of agent-based models of land cover change. In: Heppenstall AJ, Crooks AT, See LM, Batty M, eds., Agent-Based Models of Geographical Systems. Dordrecht: Springer Science+Business Media B.V., 2012.
[9] GM-VV PDG. Verification and Validation (GM-VV) to Support Acceptance of Models, Simulations and Data, GM-VV Vol. 1: Introduction and Overview. 2012; SISO-GUIDE-001.1-2012.
[10] Stainback SB, Stainback WC. Understanding and Conducting Qualitative Research. Dubuque, IA: Kendall/Hunt, 1988.